Chimeric human CstF-77/*Drosophila* Suppressor of forked proteins rescue *suppressor of forked* mutant lethality and mRNA 3′ end processing in *Drosophila*

Béatrice Benoit, François Juge, Florence Iral, Agnès Audibert*, and Martine Simonelig†

Génétique du Développement de la Drosophile, Institut de Génétique Humaine, 141 Rue de la Cardonille, 34396 Montpellier Cedex 5, France

Edited by Michael S. Levine, University of California, Berkeley, CA, and approved June 13, 2002 (received for review April 2, 2002)

The Suppressor of forked [Su(f)] protein is the *Drosophila* homologue of CstF-77, a subunit of human cleavage stimulation factor (CstF) that is required for the first step of the mRNA 3′ end processing reaction *in vitro*. We have addressed directly the role of *su(f)* in the mRNA 3′ end processing reaction *in vivo*. We show that *su(f)* is required for the cleavage of pre-mRNA during mRNA 3′ end formation. Analysis of the functional complementation between Su(f) and CstF-77 shows that most of the *Drosophila* protein (85%) can be exchanged for the human protein to produce chimeric CstF-77/Su(f) proteins that rescue lethality and cleavage defect during mRNA 3′ end formation in *su(f)* mutants. Interestingly, we show that a domain in human CstF-77 is limiting for the rescue and that this domain is not able to reproduce protein interactions with the CstF subunits of *Drosophila*. We also show that chimeric CstF-77/Su(f) proteins that rescue lethality of *su(f)* mutants cannot restore utilization of a regulated poly(A) site in *Drosophila*. Taken together, these results demonstrate that CstF-77 and Su(f) have the same function in mRNA 3′ end formation *in vivo*, but that these two proteins are not interchangeable for regulation of poly(A) site utilization.

The 3′ end processing of eukaryotic mRNA precursors (pre-mRNA) is a two-step reaction that involves endonucleolytic cleavage of the pre-mRNA and the synthesis of a poly(A) tail (1, 2). This reaction requires several protein complexes that have been purified from both yeast and mammalian cells and all or almost all components of which have been characterized. In mammals, about 15 proteins are involved and assemble into five factors that are necessary for the cleavage step of the reaction: cleavage and polyadenylation specificity factor (CPSF) (3), cleavage stimulation factor (CstF) (4), two cleavage factors (CFIm and CFIIm) (5, 6), and poly(A) polymerase (PAP) (7, 8). The second step of the reaction, polyadenylation requires CPSF, PAP, and a last protein, poly(A) binding protein II, which binds to the growing poly(A) tail and has two roles: it stimulates PAP and controls poly(A) tail length (9, 10). Assembly of the cleavage complex on the pre-mRNA results from multiple RNA–protein and protein–protein interactions. CPSF binds to a highly conserved AAUAAA element (11), the polyadenylation signal, located upstream of the cleavage site, whereas CstF recognizes a U/GU-rich element (12), located downstream of the cleavage site. Binding of CPSF and CstF to the pre-mRNA is cooperative and allows the recognition of the poly(A) site. This cooperativity results from direct protein interaction between the largest subunits (CPSF-160 and CstF-77) of these two complexes (13). In addition, binding of CFIm to the pre-mRNA appears to be an early event in the reaction and to contribute to the recruitment of the other complexes (5). CFIm was shown to interact with a subunit of CFIIm, which also interacts with CPSF, thus bridging these two complexes (6). Finally, the carboxyl-terminal domain of the largest subunit of RNA polymerase II interacts with subunits of CPSF and CstF (14) and activates the cleavage step of the reaction *in vitro* (15).

In *Drosophila*, the 3′ end processing reaction appears to be well conserved, compared with what is known in mammals, as *Drosophila* homologues are found for each of the 14 mammalian proteins shown to be involved in the reaction (16). In addition, functional conservation was demonstrated for *Drosophila* poly(A) polymerase and *Drosophila* poly(A) binding protein II in *in vitro* reconstitution assays (17, 18). Moreover, our previous data on Suppressor of forked [Su(f)], the *Drosophila* homologue of CstF-77 indicate a role for this protein in regulated utilization of poly(A) sites *in vivo* (19, 20). This finding is consistent with a role of Su(f) as part of a CstF complex in *Drosophila*.

In mammalian cells, CstF consists of three subunits of 50 kDa, 64 kDa, and 77 kDa (CstF-50, CstF-64, and CstF-77, respectively) (4). CstF-77 bridges the other two subunits of the complex (21). CstF-64 is responsible for the binding of CstF to RNA (12), as it interacts through a ribonucleoprotein-type RNA binding domain with the U/GU-rich element downstream of poly(A) sites. CstF-77 and CstF-50 interact with the carboxyl-terminal domain of RNA polymerase II (14). Data in both mammalian cells and *Drosophila* indicate a regulatory role of CstF in the utilization of alternative poly(A) sites. *In vitro*, affinity of CstF for the U/GU-rich elements, which are highly variable in sequence, defines the efficiency of poly(A) sites by determining the stability of the cleavage complex on the pre-mRNA (22). In mammalian cells, several studies have correlated shifts in the choice of poly(A) sites with quantitative or qualitative variations of CstF-64 (23–27). In *Drosophila*, we and others have shown that modulation of *su(f)* activity in *su(f)* mutants affects the utilization of alternative poly(A) sites in the f^1 mutation, the *Adh/Adhr* locus, and the *su(f)* gene itself (19, 20, 28, 29). At least for *su(f)* autoregulation, the regulated utilization of a poly(A) site by Su(f) leads to tissue-specific accumulation of a protein (20).

In this article, we have addressed more directly the role of *su(f)* in the mRNA 3′ end processing reaction *in vivo*. We show that *su(f)* is required for the cleavage of pre-mRNA during mRNA 3′ end formation and that most domains of CstF-77 and Su(f) can be interchanged *in vivo* to produce chimeric proteins that rescue lethality as well as cleavage defect in *su(f)* mutants.

Materials and Methods

***Drosophila* Stocks and Germ-Line Transformation.** The w^{1118} stock was used as a control. Homozygous $su(f)^{L26}$ mutant larvae were selected, just before they died, 72 h after egg laying at 25°C, by using the balancer chromosome *FM7c Kr-Gal4 UAS-GFPS65T* (30). We used the Gal4 driver line *P[GAL4-da.G32]* (*da-Gal4*) that mediates ubiquitous expression (31). *P* element transformation was carried out as described (32). Construct DNA (500 μg/ml) with 250 μg/ml of the helper plasmid, pUChsPΔ2–3, was

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Su(f), Suppressor of forked; CstF, cleavage stimulation factor; CPSF, cleavage and polyadenylation specificity factor; CF, cleavage factor; GST, glutathione *S*-transferase; RT-PCR, reverse transcription–PCR.

*Present address: Cycle et Détermination Cellulaire, Unité Mixte de Recherche 7622, Bât C, Boite 24, 9 Quai St Bernard, 75252 Paris Cedex 5, France.

†To whom reprint requests should be addressed. E-mail: Martine.Simonelig@igh.cnrs.fr.

injected into w^{1118} embryos. For each construct, several independent transformant lines (3 to 10) were used in the following studies.

DNA Constructs. The *UAS-CstF-77* construct was generated by cloning an *Eco*RI–*Eco*RI fragment of pG77–12 (21), which contains the *CstF-77* cDNA, into the pUAST transformation vector (33) digested with *Eco*RI. The *UAS-CstF77-NI* transgene was constructed as follows. A DNA fragment covering the 3′ coding region of *su(f)* was amplified by PCR with pcK22 (34) as a template and primers 5′-GGCGGGTCCTTTCGTCAGCGTAGAGCTACTATTCG and 5′-CCGTCTAGAGGCTTGCTTTATTGGGATTTCTTG. The amplified DNA was digested with *PpuM*I and *Xba*I, and the resulting fragment was inserted into the pG77.12 plasmid digested with *PpuM*I and *Xba*I. The inserted sequence was checked by DNA sequencing. An *Eco*RI–*Xba*I fragment of this resulting plasmid, which contains the chimeric *CstF-77*/*su(f)* cDNA, was cloned into the pUAST plasmid digested with *Eco*RI and *Xba*I, to produce the *UAS-CstF-77-NI* construct. The *UAS-CstF-77-N* transgene was constructed similarly except that PCR amplification was with primers 5′-CCACTAGTTGACCGTTACAAATTTCTGGAC and 5′-CCGTCTAGAGGCTTGCTTTATTGGGATTTCTTG and that the PCR-amplified DNA and the pG77.12 plasmid were digested with *Spe*I and *Xba*I. To construct the *UAS-CstF-77-IC* transgene, a fragment containing the 3′ region of the *CstF-77* cDNA was PCR-amplified by using pG77.12 as a template and primers 5′-CTGGTAGATAGATACAAGTTCATGG and 5′-CCGGATCCCAATACAACTTTGTTTCCAAGAACC. The amplified DNA was digested with *Pvu*II and *Bam*HI and cloned into the pJRD-*su(f)* plasmid digested with *Pvu*II and *Bam*HI. pJRD-*su(f)* had been generated as follows. The *Eco*RI–*Eco*RI fragment of pcK22 containing a full-length *su(f)* cDNA was cloned into a version of the pJRD158 plasmid in which the *Pvu*II site had been removed by digestion with *Pvu*II and *Bal*I followed by ligation. The resulting plasmid containing the *su(f)*/*CstF-77* chimeric cDNA was digested with *Eco*RI and *Bam*HI, and the *Eco*RI–*Bam*HI fragment was cloned into pUAST digested with *Eco*RI and *Bgl*II to produce the *UAS-CstF-77-IC* transgene. To generate the *UAS-CstF-77-C* transgene a 3′ coding region of *CstF-77* was PCR-amplified by using pG77.12 as a template and primers 5′-GGGCGTCTCAATCTACCAAATACTGTTGAGGAAGCTG and 5′-CCCGGTACCCAATACAACTTTGTTTCCAAGAACC. The amplified DNA was digested with *Esp3*I and *Kpn*I and cloned into the plasmid pcK22a digested with *Esp3*I and *Kpn*I. pcK22a was generated by cloning the *Eco*RI–*Eco*RI fragment of pcK22 into *Eco*RI-digested pBluescript II (Stratagene). The resulting plasmid containing the chimeric *su(f)*/*CstF-77* cDNA was digested with *Eco*RI and *Kpn*I, and the *Eco*RI–*Kpn*I fragment was cloned into pUAST to produce the *UAS-CstF-77-C* transgene.

Glutathione *S*-Transferase (GST) Pull-Down Assays. *GST-CstF-64* was obtained by cloning a *Bsp*120I–*Sca*I fragment of pZd64–19 (35) treated with Klenow into the pGEX4T2 vector digested with *Sma*I. *GST-CstF-64H* was obtained by cloning a *Mlu*I–*Eco*RV fragment of pZd64–19, filled in with Klenow, into the pGEX4T1 vector digested with *Sma*I. *GST-CstF-50K* was obtained by cloning a *Bfa*I–*Sca*I fragment of LD24780 (Berkeley *Drosophila* Genome Project), treated with Klenow, into the pGEX4T3 vector digested with *Sma*I. *GST-CstF-50KWD* was obtained by cloning the *Eco*RI–*Xho*I fragment of LD24780, filled in with Klenow, into the pGEX4T3 vector digested with *Sma*I. *GST-Su(f) 492–733* was obtained by cloning a 0.9-kb *Xho*I–*Sal*I fragment of pcK22 (34) into the pGEX4T3 vector digested with *Sal*I. GST fusion proteins were expressed and affinity-purified on glutathione-Sepharose 4B beads (Amersham Pharmacia). ^{35}S-labeled proteins were synthesized *in vitro* from DNA encoding the corresponding proteins cloned into pBluescript, by using a Promega TnT Coupled Transcription/Translation kit in the presence of [^{35}S]methionine. ^{35}S-labeled proteins (8 μl) were incubated with immobilized GST fusion proteins (5 μg) for 1.5 h at room temperature in 300 μl of Harlow buffer (50 mM Hepes, pH 7.5/100 mM NaCl/0.2 mM EDTA/1 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride/1 μg/ml leupeptine/1 mM DTT/0.5% Nonidet P-40) with 5 mg/ml of BSA. After four washes with Harlow buffer containing 200 mM NaCl, bound proteins were eluted by boiling in SDS loading buffer, resolved by SDS/PAGE, and visualized by autoradiography.

RNA Blots and Reverse Transcription–PCR (RT-PCR). RNA blots were performed as reported (36). Adult males were raised at 25°C and shifted for 4 days to 29°C. For RT-PCRs, total RNA prepared from 10-second instar larvae was dissolved in 10 μl of diethyl pyrocarbonate-treated water. One microliter of this RNA preparation was added to a 7-μl reaction containing 100 pmol random hexamers. After denaturation for 5 min at 65°C, 13 μl of a mix warmed at 42°C was added to the final concentration 1× RT buffer, 0.5 mM dNTPs, 10 mM DTT, and 10 units of Superscript-RT (GIBCO/BRL). After 1 h at 42°C, 1 μl of this reaction was put in a 50-μl PCR (30 cycles of 1 min at 94°C, 1 min at 55°C, 2 min at 72°C) with 1× PCR buffer, 1.5 mM $MgCl_2$, 50 pmol of each primer (set *rp49*: 5′-CTGCCCACCGGATTCAAGAAGT and 5′-ACTGATATCCATCCAGATAATGCA; set *sop*: 5′-GGATTGCTACACCTCGGCCCGT and 5′-CTACAACAGAATCTCCAAATCGACC; set *pgk*: 5′-GGCCAAGAAGAATAACGTGCAGTTGC and 5′-CGCTGGTCAATGCACGCACGC), 0.2 mM dNTPs, and 5 units of *Taq*DNA polymerase (Amersham Pharmacia). Five to 20 μl of the PCRs was loaded on 2% agarose gels; the same amount of the PCRs was loaded per RNA preparation. No PCR product was obtained if the RT was omitted in the reaction. RT-PCRs were performed 2–5 times with two independent RNA preparations.

Results

Rescue of *su(f)* Mutants with Chimeric CstF-77/Su(f) Proteins. Although several data support a role of *su(f)* in mRNA 3′ end formation, this function has not been demonstrated directly. To provide more direct evidence of the role of *su(f)* in mRNA 3′ end processing, we determined whether human CstF-77 has the same function *in vivo* as the Su(f) protein by testing the ability of human CstF-77 to rescue the lethality of various *su(f)* alleles. CstF-77 was expressed in *Drosophila* by using the *UAS*/*Gal4* system (33). Expression of the *CstF-77* cDNA (*UAS-CstF-77)* was driven by the *da-Gal4* activator (31), which directs ubiquitous expression of the Gal4 protein. As a control, rescue of *su(f)* mutants was assayed by using the *UAS-su(f)* transgene, which contains a full-length *su(f)* cDNA (37). The rescue of four *su(f)* mutants was analyzed. The null allele $su(f)^{L26}$ is a complete deletion of the *su(f)* locus and homozygous $su(f)^{L26}$ mutant die as larvae (38). The $su(f)^{3DES}$ allele is lethal at larval stage and it encodes a Su(f) mutant protein predicted to lack the most C-terminal 82 residues (36). The $su(f)^{R\text{-}9\text{–}18}$ allele is a ts lethal in which pupae die at 25°C and encodes a protein with a substitution at position 632 (36). The $su(f)^{ts67g}$ allele is another ts lethal, which dies at the third-instar larval stage at 29°C (39) and encodes a protein with a substitution in the N-terminal region encoded by exon 3 (K. Elliott, C. Williams, K. O'Hare, and M.S., unpublished data). The results are in Fig. 1*B*. Whereas all *UAS-su(f)* transformants were able to rescue the lethality of the four *su(f)* alleles, no rescue was obtained with the *UAS-CstF-77* lines. We next constructed chimeric CstF-77/Su(f) proteins to determine whether a particular domain of the CstF-77 protein could be responsible for the lack of rescue. Based on the known domains within Su(f) and CstF-77 (19, 21), these proteins were

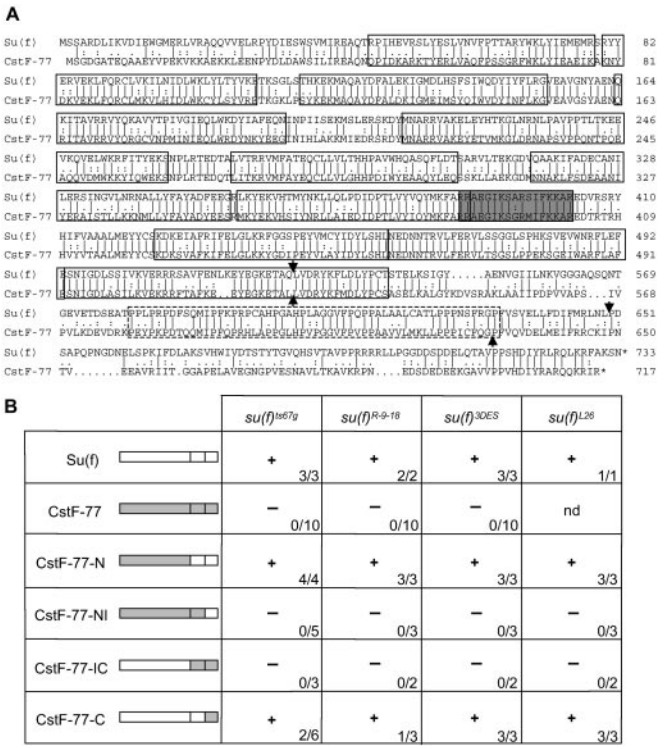

Fig. 1. Rescue of *su(f)* mutants with Su(f), CstF-77, or chimeric CstF-77/Su(f) proteins. (*A*) Comparison of Su(f) and CstF-77 proteins. Vertical lines indicate identical residues, colons represent similarities, and single dots represent less similar residues. Open boxes are tetratricopeptide-like repeats, the gray box indicates the nuclear localization signal, and the dashed box is the proline-rich domain. Arrows indicate the positions where Su(f) and CstF-77 are fused in the different chimeric proteins. (*B*) Rescue of *su(f)* mutants with Su(f), CstF-77, or chimeric CstF-77/Su(f) proteins. All of the transgenes were expressed by using the *UAS/Gal4* system with the *da-Gal4* driver. The rescue was assayed at 25°C for $su(f)^{R\text{-}9\text{-}18}$, $su(f)^{3DE5}$, and $su(f)^{L26}$ and at 29°C for $su(f)^{ts67g}$. +, Viable adults; −, lethal; nd, not determined. The numbers indicate the number of transformant lines that rescue out of the number of transformant lines tested.

separated into three regions to construct the chimeric proteins (Fig. 1*A*). The N-terminal region corresponds to the N-terminal two-thirds of the protein, which contain repeats similar to tetratricopeptide repeat motifs (19, 40). These repeats have been proposed to mediate protein–protein interactions. This region is the most conserved between Su(f) and CstF-77 (63% identity, 92% similarity), and it also contains a nuclear localization signal. The internal region contains a proline-rich domain, which is conserved between Su(f) and CstF-77 (59% identity, 81% similarity), and a region upstream of this domain, which is not conserved and may correspond to a hinge between domains of the protein (19). The C-terminal region is poorly conserved between CstF-77 and Su(f) except for a short region (28 residues) at the very C terminus. Four fusion cDNAs encoding CstF-77/Su(f) were tested for their ability to rescue the lethality of the *su(f)* mutants. A fusion protein containing the N-terminal region of CstF-77 and the other two domains from Su(f) rescues the lethality of all four *su(f)* mutants (CstF-77-N, Fig. 1*B*). Similarly, a fusion protein containing the C-terminal region of CstF-77 fused to the other two domains of Su(f) also rescues the four *su(f)* mutants (CstF-77-C, Fig. 1*B*). These results indicate that the N- and C-terminal domains that represent 85% of the protein are interchangeable *in vivo* between human and *Drosophila*. In contrast, both fusion proteins containing the internal region of CstF-77, either in addition to the CstF-77 N-terminal region or the CstF-77 C-terminal region are unable to rescue any *su(f)* mutants (CstF-77-NI and CstF-77-IC, Fig. 1*B*). This finding suggests that this internal region prevents the rescue of the *su(f)* mutants with the complete CstF-77 protein.

That 85% of the Su(f) protein can be replaced by CstF-77 to produce functional proteins in *Drosophila* indicates that Su(f) and CstF-77 have the same function *in vivo* and strongly supports a role for the Su(f) protein in mRNA 3′ end formation, as a component of a *Drosophila* CstF.

Interaction Between Human and *Drosophila* Subunits of CstF. Protein interactions within mammalian CstF have been analyzed *in vitro* (41). CstF-77 was shown to interact with itself and the other two subunits of the complex, and the domain responsible for these interactions was identified as the internal region containing the proline-rich domain and the short hinge region upstream of it (41). The proline-rich domain alone was shown to interact with CstF-64 and the proline-rich domain in addition to the hinge region was shown to mediate interaction with CstF-50 and CstF-77. We found that the region of CstF-77 that is limiting for the rescue of *su(f)* mutant lethality is the same internal region that is responsible for protein interactions within human CstF. We, therefore, reasoned that the lack of rescue with chimeric proteins containing this internal domain from the human protein could result from altered interactions within CstF. To test this hypothesis, we performed GST pull-down assays to analyze interactions between *Drosophila* homologues of CstF subunits and interactions between these homologues and human CstF-77 or chimeric CstF-77/Su(f) proteins. GST fusion proteins were generated from Su(f), *Drosophila* CstF-64, and *Drosophila* CstF-50. The domains responsible for interaction within CstF were determined in human CstF-64 and CstF-50 proteins *in vitro* (41). The domain in CstF-64 that mediates interaction with CstF-77 was shown to be a domain called "hinge," located between a ribonucleoprotein-type RNA binding domain and a long proline/glycine-rich region. Human CstF-50 interacts with both CstF-77 and itself. Self-interaction occurs through the N-terminal region of the protein, whereas interaction with CstF-77 appears to require the seven WD-40 repeats. We generated GST fusion proteins containing either the regions known to be involved in CstF interactions or longer versions of the protein for *Drosophila* CstF-64 and CstF-50 (Fig. 2*A*). The capacity of these GST fusion proteins immobilized on glutathione-Sepharose beads to interact with ^{35}S-labeled complete Su(f), human CstF-77, or chimeric CstF-77/Su(f) proteins synthesized *in vitro* was analyzed. Fig. 2*B* shows that interactions within CstF are conserved in *Drosophila*. Complete Su(f) binds to a C-terminal region of Su(f) that contains the hinge and proline-rich domains. It also binds to *Drosophila* CstF-64 and CstF-50; in both cases however, the shorter versions of these proteins (CstF-64H and CstF-50WD) interact less efficiently with complete Su(f) than the longer versions (Fig. 2*B*, Su(f), compare lanes 4 to 5 and 6 to 7). Similar results were obtained with the CstF-77/Su(f) chimeric proteins that rescue *su(f)* mutant lethality *in vivo*, CstF-77-N and CstF-77-C (Fig. 2*B*). These two chimeric proteins also bind to the C-terminal region of Su(f) and to *Drosophila* CstF-64 and CstF-50 with an efficiency comparable to that of complete Su(f). In contrast, dramatically different results were obtained when interactions between human CstF-77 and *Drosophila* CstF components were analyzed (Fig. 2*B*). Human CstF-77 does not bind to the C-terminal region of Su(f), nor does it bind to *Drosophila* CstF-50 (Fig. 2*B*, CstF-77, lanes 3, 6, and 7). Moreover, human CstF-77 binds to *Drosophila* CstF-64 and this interaction is clearly stronger than that between Su(f) and *Drosophila* CstF-64 (Fig. 2*B*, lanes 4 and 5, compare Su(f) to CstF-77). Interactions with the chimeric protein CstF-77-NI that contains the hinge and proline-rich domains from human CstF-77 and that does not rescue *su(f)* mutants *in vivo* were found to be similar to that of human CstF-77. This chimeric protein

GENETICS

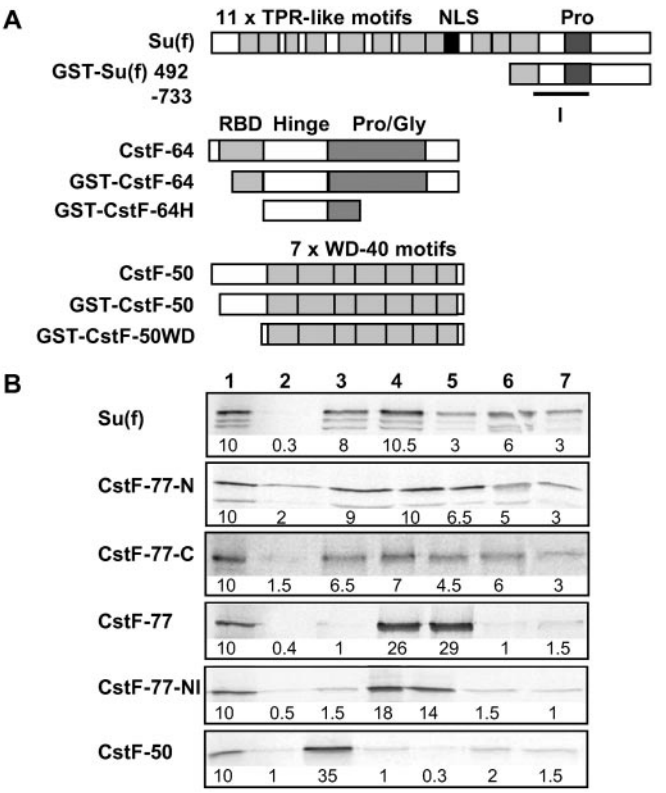

Fig. 2. Protein interactions within *Drosophila* CstF and between human CstF-77 and *Drosophila* CstF subunits. (*A*) Schematic representation of *Drosophila* CstF subunits and GST fusion proteins. For each GST fusion protein, the part of the *Drosophila* protein fused to GST is indicated. The internal domain of Su(f), containing the proline-rich (Pro) and hinge domains, that was exchanged in chimeric CstF-77/Su(f) proteins is indicated (I). TPR, tetratricopeptide repeats; NLS, nuclear localization signal; RBD, RNA binding domain. (*B*) GST pull-down assays. ^{35}S-labeled Su(f), CstF-77, chimeric CstF-77/Su(f), or *Drosophila* CstF-50 proteins were incubated with GST or GST fusion proteins. Incubations were in the presence of 40 μl of glutathione-Sepharose beads. Labeled proteins eluted from the beads, as well as 10% of the input-labeled proteins (input), were analyzed by electrophoresis and quantified with IMAGE QUANT. Amounts are indicated in percentages, relative to the input. Lane 1, input; lane 2, GST; lane 3, GST-Su(f)492–733; lane 4, GST-CstF-64; lane 5, GST-CstF-64H; lane 6, GST-CstF-50; lane 7, GST-CstF-50WD.

does not interact with Su(f) or *Drosophila* CstF-50, but it strongly interacts with *Drosophila* CstF-64. As a negative control of interaction with CstF-64, we assayed binding of ^{35}S-labeled complete *Drosophila* CstF-50 to the GST fusion proteins. As expected, *Drosophila* CstF-50 binds to the C-terminal part of Su(f) that includes the hinge and proline-rich domains, but does not bind to *Drosophila* CstF-64 (Fig. 2*B*, CstF-50). In this experiment, self-interaction of CstF-50 is not detected or extremely weak. This result could come from the truncation in the GST fusion protein of the N-terminal domain of CstF-50 (Fig. 2*A*), known to be involved in self-interaction in the human protein (41).

Analysis of these protein interactions *in vitro* provides an explanation of why human CstF-77 and chimeric CstF-77/Su(f) proteins containing the hinge and proline-rich domains from the human protein are not able to replace Su(f) in *Drosophila*. When present in the place of Su(f) in *Drosophila*, human CstF-77 or chimeric CstF-77-NI and CstF-77-IC should be able to interact with themselves and *Drosophila* CstF-64, but not with *Drosophila* CstF-50. Therefore, it is probable that a so-called CstF complex does not exist in this context.

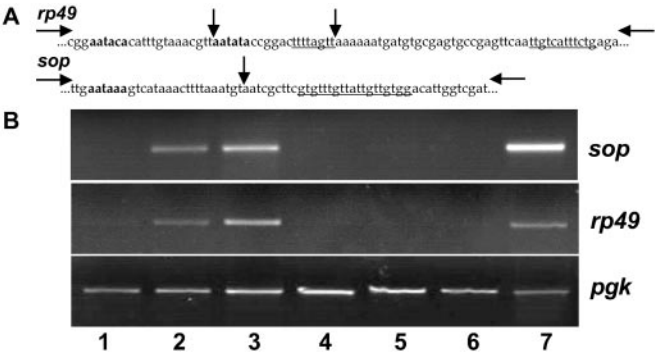

Fig. 3. Rescue of the cleavage defect in the *su(f)^{L26}* mutant by chimeric CstF-77/Su(f) proteins. (*A*) Sequences of *rp49* and *sop* poly(A) site regions. Potential poly(A) signals are in bold. Vertical arrows indicate poly(A) sites determined from cDNA sequences in databases. GU/U-rich sequences are underlined. Horizontal arrows indicate primers used for the PCR. (*B*) RT-PCR assays. The control *pgk* PCR fragment is generated with primers on each side of intron 2 of *pgk*. The size of the *pgk* PCR product we obtained (434 bp) was the expected size for amplification of *pgk* RNA after splicing of intron 2. *rp49* (369 bp) and *sop* (316 bp) PCR fragments are obtained with primers indicated in *A*, only if no cleavage occurs at the poly(A) site. Lane 1, wild type; lane 2, *su(f)^{L26}*; lane 3, *su(f)^{L26};da-Gal4*; lane 4, *su(f)^{L26};UAS-CstF-77-C/+;da-Gal4/+*; lane 5, *su(f)^{L26};UAS-CstF-77-N/da-Gal4*; lane 6, *su(f)^{L26};UAS-su(f)/da-Gal4*; lane 7, *su(f)^{L26};UAS-CstF-77/da-Gal4*.

Rescue of Cleavage Defect in *su(f)* Mutants by Chimeric CstF-77/Su(f) Proteins. We wanted to know whether chimeric CstF-77/Su(f) proteins that rescue *su(f)* mutant lethality *in vivo* have indeed the same role in mRNA 3′ end processing, as has been determined for human CstF-77 *in vitro* (21). We, therefore, developed an assay to measure cleavage during the mRNA 3′ end processing reaction *in vivo*. This assay consists of looking by RT-PCR for RNA molecules that have been uncleaved at a poly(A) site. We used the two genes, *rp49* and *sop*, that encode ribosomal proteins, because they are highly and ubiquitously expressed (42, 43). For each gene, primers for the PCR are selected on each side of the poly(A) site (Fig. 3*A*), such that if cleavage occurs normally, no PCR product or a very low amount of PCR product is expected. Total RNA was prepared from wild-type and *su(f)^{L26}* larvae. RNA was controlled by a RT-PCR with primers located in the coding region, on both sides of an intron of the *pgk* gene, another gene expressed ubiquitously (44). Fig. 3*B* shows that in the wild type cleavage occurs at the poly(A) sites of *sop* and *rp49*, as no or a very low amount of PCR product is detected (Fig. 3*B*, lane 1). In contrast, in *su(f)^{L26}* mutant larvae, PCR products were amplified for both *rp49* and *sop* (Fig. 3*B*, lanes 2 and 3). This result indicates that in a null allele of *su(f)*, uncleaved pre-mRNAs that contain sequences downstream of poly(A) sites accumulate. These data strongly support a role of the Su(f) protein in the cleavage step of the mRNA 3′ end processing reaction *in vivo*.

We tried to rescue this cleavage defect with Su(f) and chimeric CstF-77/Su(f) proteins. When *UAS-su(f)* is expressed ubiquitously in *su(f)^{L26}* mutant larvae, cleavage of pre-mRNAs at poly(A) sites is restored, as no PCR product is detected for both *rp49* and *sop* (Fig. 3*B*, lane 6). The same result was obtained with chimeric CstF-77/Su(f) proteins that rescue *su(f)* mutant lethality, CstF-77-N and CstF-77-C. When these chimeric proteins are expressed ubiquitously in *su(f)^{L26}* larvae, no PCR product overlapping the poly(A) site of *rp49* and *sop* is detected (Fig. 3*B*, lanes 4 and 5). As a control, we used the complete human CstF-77 protein that does not rescue *su(f)* mutants. Expression of this protein in *su(f)^{L26}* larvae does not restore cleavage at *rp49* and *sop* poly(A) sites as uncleaved pre-mRNAs can be amplified by the RT-PCR (Fig. 3*B*, lane 7).

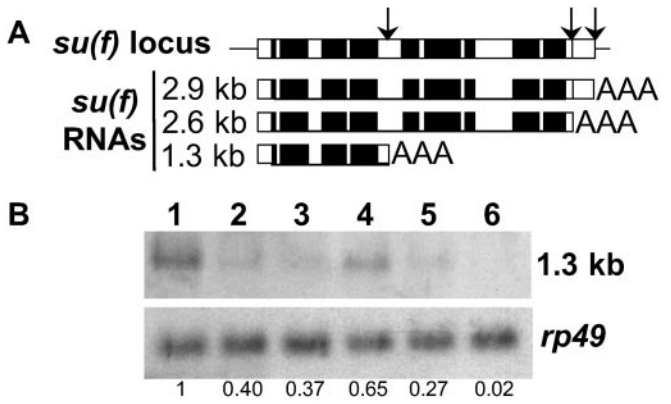

Fig. 4. Utilization of the *su(f)* intronic poly(A) site in the $su(f)^{ts67g}$ mutant is not restored with the chimeric CstF-77/Su(f) proteins. (*A*) Structure of the *su(f)* locus and RNAs. Open boxes are noncoding sequences and introns, black boxes are coding sequences. Vertical arrows indicate poly(A) sites. (*B*) RNA blots with poly(A)$^+$ RNA from wild-type and $su(f)^{ts67g}$ adult males raised at 25°C and shifted for 4 days to 29°C were hybridized with a *su(f)* RNA probe specific to the 1.3-kb RNA. This probe is complementary to 88 nt located in the part of intron 4 incorporated in the 1.3-kb RNA. The blots were reprobed with the *rp49* clone as a loading control. Quantification was performed with National Institutes of Health IMAGE. The 1.3-kb/*rp49* ratios are indicated relative to the value obtained for the wild type, set at 1. Lane 1, wild type; lane 2, $su(f)^{ts67g}$; lane 3, $su(f)^{ts67g}$*;da-Gal4*; lane 4, $su(f)^{ts67g}$*;UAS-su(f)/da-Gal4*; lane 5, $su(f)^{ts67g}$*;UAS-CstF-77-N/da-Gal4*; lane 6, $su(f)^{ts67g}$*;UAS-CstF-77-C/+; da-Gal4/+*.

That CstF-77/Su(f) chimeric proteins, which contain either 72% of human CstF-77 at the N terminus or 13% of the human protein at the C terminus, restore mRNA 3′ end cleavage in a *su(f)* mutant demonstrate that Su(f) and human CstF-77 have the same molecular function *in vivo* in the cleavage step of the mRNA 3′ end processing reaction.

Utilization of a Regulated Poly(A) Site in *Drosophila* Is Not Restored by Chimeric CstF-77/Su(f) Proteins. To further confirm the identity of function between Su(f) and CstF-77, we wanted to test whether the CstF-77/Su(f) fusion proteins are also able to replace the Su(f) protein in the utilization of a regulated poly(A) site. The *su(f)* gene has three alternative poly(A) sites, one of which is located in an intron (34) (Fig. 4*A*). We have shown previously that utilization of this *su(f)* intronic poly(A) site is particularly sensitive to *su(f)* activity. Its utilization leads to the formation of a truncated 1.3-kb transcript, the amount of which decreases in *su(f)* mutants compared with wild type (19). We have also shown that accumulation of this transcript is tissue specific, as it is present in a low amount in mitotic cells and a higher amount in nonmitotic cells (20). This results in the tissue-specific expression pattern of the Su(f) protein and depends on regulated utilization of the *su(f)* intronic poly(A) site at least in part by the Su(f) protein itself (20). We tested the ability of the chimeric proteins CstF-77-N and CstF-77-C that rescue *su(f)* mutant lethality and utilization of poly(A) sites in *sop* and *rp49* to promote the utilization of this *su(f)* regulated poly(A) site by using Northern blots. In the $su(f)^{ts67g}$ mutant, at the restrictive temperature (29°C), the amount of the 1.3-kb *su(f)* truncated transcript is low (Fig. 4*B*, lanes 2 and 3), compared with wild type (Fig. 4*B*, lane 1). As expected, when the Su(f) protein is provided ubiquitously in $su(f)^{ts67g}$ individuals at the restrictive temperature, an increased amount of the 1.3-kb truncated transcript is observed (Fig. 4*B*, lane 4) (20). In contrast, such an increase is not detected when CstF-77-N or CstF-77-C chimeric proteins are expressed in the $su(f)^{ts67g}$ mutant (Fig. 4*B*, lanes 5 and 6). This result shows that these two chimeric proteins, although able to replace Su(f) for the utilization of unregulated poly(A) sites, are unable to promote utilization of a regulated poly(A) site in *Drosophila*.

Discussion

CstF and its Function Are Conserved in *Drosophila*. Data in this article demonstrate that, *in vivo*, *su(f)* has the function in mRNA 3′ end processing that is predicted from *in vitro* studies with its human homologue CstF-77 (21). We show that the Su(f) protein is required for the cleavage of pre-mRNA at unregulated poly(A) sites. In addition, we show that poly(A) site cleavage defect in a *su(f)* null mutant is rescued by chimeric CstF-77/Su(f) proteins. These chimeric CstF-77/Su(f) proteins are also able to rescue the lethality of *su(f)* mutants. Using these two assays, we find that the complete protein except for an internal short region (15% of the protein) can be interchanged between *Drosophila* and human. These results show that lethality in *su(f)* mutants results from the defect in poly(A) site utilization and that Su(f) has the same function in pre-mRNA cleavage at poly(A) sites as its human homologue CstF-77.

Analysis of protein interactions within *Drosophila* CstF also shows that this complex is conserved in *Drosophila*. Human CstF-77 was shown to interact *in vitro* with itself, CstF-64, and CstF-50, and CstF-50 was also shown to interact with itself (41). Results of GST pull-down assays show that the Su(f) protein interacts *in vitro* with *Drosophila* CstF-64, CstF-50, and itself. As in mammals, we found that the hinge domain of *Drosophila* CstF-64 is sufficient for interaction with Su(f). However, a larger version of CstF-64 has a better affinity for Su(f) than the hinge domain alone, suggesting that the rest of the protein may also contribute to this binding. Similarly, the WD-40 repeats of *Drosophila* CstF-50 are sufficient for interaction with Su(f), but the strength of this interaction is increased when a near complete CstF-50 is used. We also confirmed that an internal domain of Su(f) containing the proline-rich domain and the nonconserved region upstream of it (hinge domain) is involved in self-association and in interaction with CstF-50 as these interactions are lost when this internal domain is changed in Su(f).

Self-association of CstF-77 and CstF-50 *in vitro* indicates that the CstF complex could dimerize. Data in *Drosophila* suggest that this could occur *in vivo*. In *Drosophila*, complementation to viability between lethal alleles of *su(f)* that affect the Su(f) protein in different domains has led us to propose a Su(f)–Su(f) interaction (36). The fact that viability is restored by interaction between mutant Su(f) proteins could indicate that CstF dimerizes at some step during the polyadenylation reaction.

We show that the hinge and proline-rich domains of human CstF-77 cannot replace corresponding domains of the *Drosophila* protein for protein interactions within CstF. Human CstF-77 or a chimeric protein with the internal region (hinge and proline-rich domains) from the human protein interact strongly with *Drosophila* CstF-64 and do not interact with *Drosophila* CstF-50 and Su(f). In contrast, Su(f) or chimeric proteins with this internal region from Su(f) interact with all three *Drosophila* CstF proteins. We propose that this change is responsible for the lack of rescue of *su(f)* mutants with CstF-77 and chimeric proteins containing this internal region from CstF-77. When one of these proteins replaces Su(f) in *Drosophila*, a CstF complex probably does not form.

Regulated Protein Interactions Within CstF. With human proteins, *in vitro* interaction between CstF-64 and a truncated form of CstF-77, containing the proline-rich domain and a part of the hinge domain, is stronger than that observed with full-length CstF-77 (41). Interestingly, this interaction decreases to wild-type level if the other part of the hinge domain and the last tetratricopeptide repeat-like domain of CstF-77 are added. This finding suggests that the CstF-77 hinge domain could have a

function in regulating affinity of CstF-77 for CstF-64. The sequence of the hinge domain is not conserved between human and *Drosophila* (Fig. 1*A*), and we found that human CstF-77 hinge and proline-rich domains have a high affinity for *Drosophila* CstF-64. In contrast, interaction between *Drosophila* CstF-64 and Su(f) is weaker. Together, these data suggest that there is a selection pressure, in both species, to maintain a weak interaction between CstF-64 and CstF-77. As sequence requirements in CstF-77 for interaction with the three subunits of the complex include the same proline-rich domain, it is possible that a strong interaction with CstF-64 would prevent interaction with CstF-50 and/or CstF-77, thus precluding the formation or activity of the CstF complex.

CstF and the Regulation of Alternative Poly(A) Site Choice in Human and *Drosophila*. In human as in *Drosophila,* CstF activity appears to be essential for poly(A) site choice. In human, a shift in poly(A) site choice in the IgM heavy-chain gene occurs during B cell maturation. An increase in CstF-64 protein amount or activity during differentiation has been proposed to be involved in utilization of a weak proximal poly(A) site of the gene, which generates a transcript coding for a secreted protein in mature B cells (24, 25). In *Drosophila,* we previously showed that the Su(f) protein is required for tissue-specific utilization of a regulated poly(A) site in intron 4 of the *su(f)* gene (19, 20). We tested the ability of chimeric CstF-77/Su(f) proteins that rescue *su(f)* mutant lethality to also rescue the utilization of the *su(f)* regulated intronic poly(A) site *in vivo*. These two chimeric proteins, CstF-77-N and CstF-77-C, are able to restore utilization of nonregulated poly(A) sites (Fig. 3); however, they do not promote the utilization of the *su(f)* intronic poly(A) site. This poly(A) site has a downstream GU-rich sequence, but it is not preceded by the consensus AAUAAA element and was proposed earlier to be a weak poly(A) site (19). Taken together, these data indicate that the chimeric proteins CstF-77-N and CstF-77-C are active in mRNA 3′ end processing, but are not able to induce utilization of a weak regulated poly(A) site. These results suggest that Su(f) activity must be optimal for regulated utilization of poly(A) sites and correlate with studies of weak *su(f)* mutants, in which utilization of regulated poly(A) sites are affected only (19, 20, 29). In the chimeric CstF-77-N protein, the human N-terminal domain is limiting for utilization of the *su(f)* intronic poly(A) site. This domain contains tetratricopeptide repeat-like motifs and is supposed to be involved in binding to CPSF-160 (41). We can therefore hypothesize that wild-type interaction between CPSF and CstF is essential for utilization of regulated poly(A) sites. Interaction between human N-terminal domain of CstF-77 (in CstF-77-N) and *Drosophila* CPSF-160 would not allow the formation of a cleavage complex stable enough on a weak regulated poly(A) site. The C-terminal domain of CstF-77/Su(f) has not been proposed to mediate protein interactions, but the fact that CstF-77-C does not induce utilization of the *su(f)* intronic poly(A) site suggests that this C-terminal domain is important for regulated utilization of weak poly(A) sites *in vivo*. This C-terminal region of Su(f) is necessary for *su(f)* function *in vivo*, as deletions of this domain lead to *su(f)* lethal alleles (36). It is possible that this C-terminal domain interacts with particular proteins that could potentially regulate CstF activity and poly(A) site choice.

This work was initiated in the laboratory Dynamique du Génome et Evolution, Institut Jacques Monod, Paris, and we are grateful to Prof. D. Anxolabéhère for support during this period. We thank Dr. K. O'Hare for the gift of pG77.12 and pZd64–19 plasmids. We are very grateful to Prof. E. Wahle for helpful discussions. This work was supported by the Centre National de la Recherche Scientifique (Unité Mixte de Recherche 9922, Unité Propre de Recherche 1142), the Universities P. M. Curie, D. Diderot, the Association pour la Recherche sur le Cancer (No. 9885) and the Fondation pour la Recherche Médicale. B.B., F.J., and A.A. held awards from the Ministère de l'Enseignement Supérieur et de la Recherche, the Association pour la Recherche sur le Cancer, and the Fondation pour la Recherche Médicale.

1. Wahle, E. & Rüegsegger, U. (1999) *FEMS Microbiol. Rev.* **23,** 277–295.
2. Zhao, J., Hyman, L. & Moore, C. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 405–445.
3. Bienroth, S., Wahle, E., Suter-Crazzolara, C. & Keller, W. (1991) *J. Biol. Chem.* **266,** 19768–19776.
4. Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J. & Shenk, T. (1990) *Genes Dev.* **4,** 2112–2120.
5. Rüegsegger, U., Blank, D. & Keller, W. (1998) *Mol. Cell* **1,** 243–253.
6. de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H. & Keller, W. (2000) *EMBO J.* **19,** 5895–5904.
7. Raabe, T., Bollum, F. J. & Manley, J. L. (1991) *Nature (London)* **353,** 229–234.
8. Wahle, E., Martin, G., Schiltz, E. & Keller, W. (1991) *EMBO J.* **10,** 4251–4257.
9. Bienroth, S., Keller, W. & Wahle, E. (1993) *EMBO J.* **12,** 585–594.
10. Wahle, E. (1995) *J. Biol. Chem.* **270,** 2800–2808.
11. Keller, W., Bienroth, S., Lang, K. M. & Christofori, G. (1991) *EMBO J.* **10,** 4241–4249.
12. MacDonald, C. C., Wilusz, J. & Shenk, T. (1994) *Mol. Cell. Biol.* **14,** 6647–6654.
13. Murthy, K. G. K. & Manley, J. L. (1995) *Genes Dev.* **9,** 2672–2683.
14. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M. & Bentley, D. L. (1997) *Nature (London)* **385,** 357–361.
15. Hirose, Y. & Manley, J. L. (1998) *Nature (London)* **395,** 93–96.
16. Mount, S. M. & Salz, H. K. (2000) *J. Cell. Biol.* **150,** F37–F44.
17. Benoit, B., Nemeth, A., Aulner, N., Kühn, U., Simonelig, M., Wahle, E. & Bourbon, H. M. (1999) *Nucleic Acids Res.* **27,** 3771–3778.
18. Murata, T., Nagaso, H., Kashiwabara, S., Baba, T., Okano, H. & Yokoyama, K. K. (2001) *Dev. Biol.* **233,** 137–147.
19. Audibert, A. & Simonelig, M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14302–14307.
20. Juge, F., Audibert, A., Benoit, B. & Simonelig, M. (2000) *RNA* **6,** 1529–1538.
21. Takagaki, Y. & Manley, J. L. (1994) *Nature (London)* **372,** 471–474.
22. Weiss, E. A., Gilmartin, G. M. & Nevins, J. R. (1991) *EMBO J.* **10,** 215–219.
23. Mann, K. P., Weiss, E. A. & Nevins, J. R. (1993) *Mol. Cell. Biol.* **13,** 2411–2419.
24. Edwalds-Gilbert, G. & Milcarek, C. (1995) *Mol. Cell. Biol.* **15,** 6420–6429.
25. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. (1996) *Cell* **87,** 941–952.
26. Takagaki, Y. & Manley, J. L. (1998) *Mol. Cell* **2,** 761–771.
27. Veraldi, K. L., Arhin, G. K., Martincic, K., Chung-Ganster, L. H., Wilusz, J. & Milcarek, C. (2001) *Mol. Cell. Biol.* **21,** 1228–1238.
28. Ishimaru, S. & Saigo, K. (1993) *Mol. Gen. Genet.* **241,** 647–656.
29. Brogna, S. & Ashburner, M. (1997) *EMBO J.* **16,** 2023–2031.
30. Casso, D., Ramirez-Weber, F. A. & Kornberg, T. B. (1999) *Mech. Dev.* **88,** 229–232.
31. Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. (1995) *Cell* **82,** 67–76.
32. Rubin, G. M. & Spradling, A. C. (1982) *Science* **218,** 348–353.
33. Brand, A. H. & Perrimon, N. (1993) *Development (Cambridge, U.K.)* **118,** 401–415.
34. Mitchelson, A., Simonelig, M., Williams, C. & O'Hare, K. (1993) *Genes Dev.* **7,** 241–249.
35. Hatton, L. S., Eloranta, J. J., Figueiredo, L. M., Takagaki, Y., Manley, J. M. & O'Hare, K. (2000) *Nucleic Acids Res.* **28,** 520–526.
36. Simonelig, M., Elliott, K., Mitchelson, A. & O'Hare, K. (1996) *Genetics* **142,** 1225–1235.
37. Audibert, A., Juge, F. & Simonelig, M. (1998) *Mech. Dev.* **72,** 53–63.
38. Audibert, A. & Simonelig, M. (1999) *Mech. Dev.* **82,** 41–50.
39. Dudick, M. E., Wright, T. R. F. & Brothers, L. L. (1974) *Genetics* **76,** 487–510.
40. Preker, P. J. & Keller, W. (1998) *Trends Biochem. Sci.* **23,** 15–16.
41. Takagaki, Y. & Manley, J. L. (2000) *Mol. Cell. Biol.* **20,** 1515–1525.
42. O'Connell, P. O. & Rosbash, M. (1984) *Nucleic Acids Res.* **12,** 5495–5513.
43. Barrio, R., del Arco, A., Cabrera, H. L. & Arribas, C. (1993) *Nucleic Acids Res.* **21,** 351.
44. Roselli-Rehfuss, L., Ye, F., Lissemore, J. L. & Sullivan, D. T. (1992) *Mol. Gen. Genet.* **235,** 213–220.